The EuroAsiaSpeller version 6.2.3.7/W (RC1)
UNICODE
User manual

*TALŌ b.v.,
Lijsterlaan 379,
1403 AZ Bussum, NL.

Spelling of lexical and grammatical collocations
Enlarged edition
Completely revised

*TALŌ

the Germanic root of our words *tell*, *tale*, like a *notch made on a tally*

May 2015

Copyright © *TALŌ b.v., 2001, ... ,2015

All rights reserved. Without limiting the rights under copyright reserved above, no part of this production may be reproduced, stored in or introduced into a retrieval system or transmitted, in any form or by any means (electronic, mechanical, photocopying, recording or otherwise), without the prior written permission of both the copyright owner and the above publisher of this book

The greatest care has been taken in compiling this book. However, no responsibility can be accepted by the publisher or author for the accuracy of the information presented.

1. **INTRODUCTION**

The EuroAsiaSpeller adds spelling capabilities to text editors for the European, Asian, African, North American languages.

The EuroAsiaSpeller is a single library speller system supporting nearly any European languages and beyond.

The EuroAsiaSpeller presents new standardized method to implement spelling.

The EuroAsiaSpeller adds control of punctuation.

The EuroAsiaSpeller adds control of lexical and grammatical collocations.

The EuroAsiaSpeller comes with one text tool: EUSpell (an ascii text editor).

The EuroAsiaSpeller adds correction of expressions enabling, and support many languages using complex scripts such as Hindi, Marathi, Nepalese, Sinhala. *your own bespoken style guide.*

The EuroAsiaSpeller DLL/Shared Library can also be attached to other publishing systems. In order to be successful you will need the Shared Library's technical documentation "talo_s_lib6236.pdf" (information: info@talo.nl) and the library. An introduction can be found in the directory "EuroAsiaSpeller/euspell/doc/technotes".

1.1. **Distribution**

The version 6.2.3 of the EuroAsiaSpeller is distributed for OEM-clients on a CD- or DVD-medium or is available from Internet ("http://www.talo.nl/" download menu).

- The CD combines all lexicons on a single medium.
- The Internet extracts of the CD are divided into multiple zip-files: the main utility *euspell_main_6.2.3.zip* and several sets of dictionaries, e.g., *euspell_lex_uk-am-ca_6.2.3.zip* or *euspell_lex_de_6.2.3.zip*, the English and German lexicons. To install from Internet at least two modules should be downloaded and unpacked (the unpack path has to be the same).

1.2. **Try before you buy!**

You can try the Internet extracts before buying, for a period of one month per language module! You cannot install a language for a second time.
Please inquire about any information *how to order* the TALO's spell checker at *e-mail: info@talo.nl*.
If you do not buy a license, you have to erase (or uninstall) the program and accompanying files on all storage media.

2. **Key features**

- Spelling reforms, automatic respelling
- Accurate suggestions
- Add your own alternatives to user mistakes instead of ours
- Separate Dialogues for Alternatives and Loop Up for Words
- Fast access to any dictionary entry independent of the size
- Renew integrated punctuation checks during spelling
- Integrated checks for lexical and grammatical collocations during spelling which can serve as a newspaper's *Style Guide*
- Learning from history, automatic correction
- Tunable Accuracy, text and OCR modes
- Three button Concept for a comprehensible set of spelling functions
- UNICODE (UTF-16 and UTF-8) for Windows XP, Vista, Windows 7 & 8
- DialogLess Shared Libraries for Linux x86 and Linux x64 available
- Sorting Order either alphabetical or based on similarity
- Apply User Actions Session Only or Add Change Warning to User Actions at time of Spell Checking
- A Stop Words list to skip unsuitable suggestions in spell check dialogs

3. **INSTALLATION**

This installation program is automatically executed when you put the CD in the CD-drive or when the CD is entered in the CD drive. Alternatively the user can execute the [CD]:setup.exe or setup_x64 program, which executes the installation program [CD]:\euspell\bin32\installs.exe (Note, this complete path has always to be used). The install procedure invites you to supply a general license key and a license key per language (group) to be specified during installation of the dictionaries. For Windows 32 bits versions "setup.exe" should be used and NOT "setup_x64.exe". "Setup_x64.exe" is meant for 64 bit versions of Windows ONLY. For Windows Vista, Windows 7 and Windows 8 the setup program should be **"run as administrator"**, however in most cases the setup utility will probably be run as administrator by default. If not choose the above option yourself. In addition you should allow the Setup Utility to run!
For a demo installation the general license key should be set to zero or better the area of license key should left blank. For all other cases key codes are required.

The Internet extracts use a virtual CD, a path name which replaces the CD. This path name is c:\EuroAsiaSpeller by default. The path name is automatically altered during the installation and only a few clicks on <Next>, <Yes> or <OK> buttons are required.
The the EuroAsiaSpeller SetUp procedure starts with a welcome screen, as presented in fig. 1. By clicking on Next the program continues, by clicking on Exit the program will close.

By clicking on Next the license agreement is shown, which can be accepted (Yes) or rejected (No). If you don't accept the license agreement click on <No>. SetUp will close. If you accept the license agreement click on <Yes>. SetUp will continue.
If you have accepted the license agreement SetUp continues and the user is guided though the installation procedure by a few of steps that acquire simple actions.
During the SetUp an explicative message log is written to the background screen. PLEASE, READ THE TEXT on the background log screen. You can move the windows in top. The first step is to supply the speller with basic information (see fig. 3).
If you install the "EuroAsiaSpeller", please, type your authorization number (General Access Key) and the other registration variables in the edit fields of the EuroAsiaSpeller Authorization dialogue. The authorization number consists of a series of numbers separated by a hyphen.
For the option "try before you buy" or if you don't have an authorization

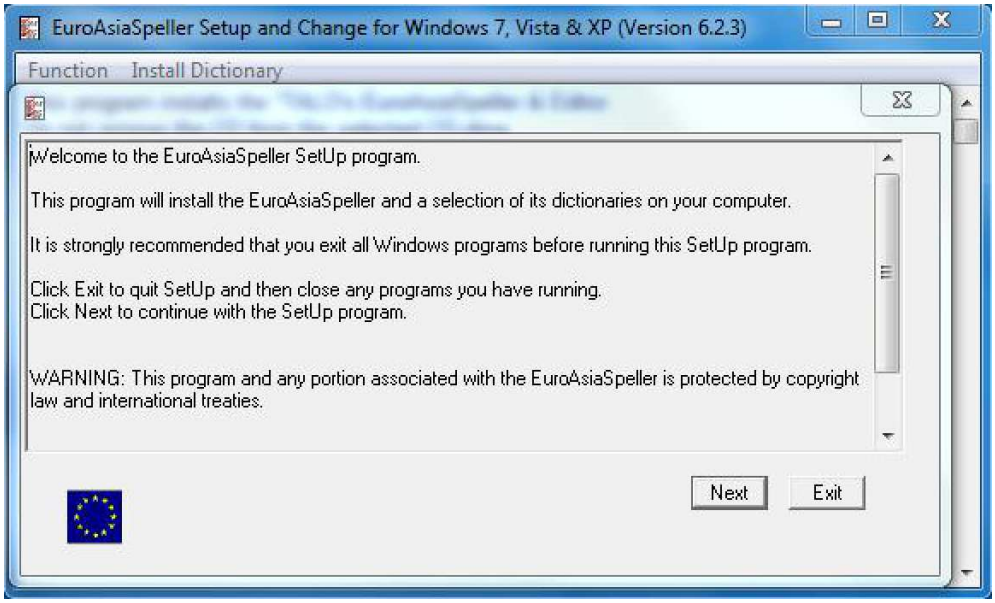
EuroAsiaSpeller Setup and Change for Windows 7, Vista & XP (Version 6.2.3)
Function Install Dictionary
Welcome to the EuroAsiaSpeller SetUp program.
This program will install the EuroAsiaSpeller and a selection of its dictionaries on your computer.
It is strongly recommended that you exit all Windows programs before running this SetUp program.
Click Exit to quit SetUp and then close any programs you have running.
Click Next to continue with the SetUp program.
WARNING: This program and any portion associated with the EuroAsiaSpeller is protected by copyright law and international treaties.
Next
Exit

Fig. 1: The welcome screen of the EuroAsiaSpeller SetUp. For continue click <Next>, for close click <Exit>.

number yet and are testing the EuroAsiaSpeller you should leave the Authorization field blank.

Type your Authorization number: xxxx-xxxx-xxxx
Type your name for registering: my_name
Type your company name: my_company_name
and press <Continue>

If you want to uninstall the EuroAsiaSpeller, please, press the <**Uninstall**> Button and follow the instructions.

After having pressed <Next> the installation continues with the SetUp parameters.

The "EuroAsiaSpeller Set Up Parameters" dialogue box appears above the background log screen.

By default the CD-station and four directories are displayed in the Set Up Dialogue:

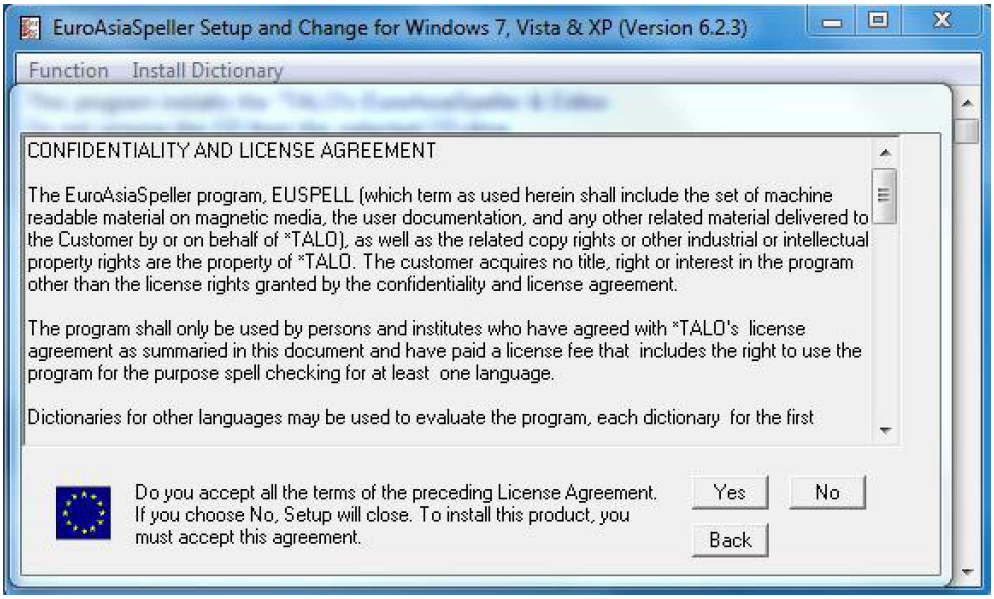
EuroAsiaSpeller Setup and Change for Windows 7, Vista & XP (Version 6.2.3)
Function Install Dictionary
CONFIDENTIALITY AND LICENSE AGREEMENT
The EuroAsiaSpeller program, EUSPELL (which term as used herein shall include the set of machine readable material on magnetic media, the user documentation, and any other related material delivered to the Customer by or on behalf of *TALO), as well as the related copy rights or other industrial or intellectual property rights are the property of *TALO. The customer acquires no title, right or interest in the program other than the license rights granted by the confidentiality and license agreement.
The program shall only be used by persons and institutes who have agreed with *TALO's license agreement as summaried in this document and have paid a license fee that includes the right to use the program for the purpose spell checking for at least one language.
Dictionaries for other languages may be used to evaluate the program, each dictionary for the first
Do you accept all the terms of the preceding License Agreement. If you choose No, Setup will close. To install this product, you must accept this agreement.
Yes
No
Back

Fig. 2: The license agreement text, which has to be accepted to continue.

Windows XP
d:
c:\Program Files\euspell\bin32 (program directory)
c:\Program Files\euspell\spell (lexicon directory)
c:\Program Files\euspell\spell (server directory)
c:\Program Files\euspell\spell (user history or personal directory)
c:\tmp
Windows Vista, Windows 7 and Windows 8
d:
c:\Users\Public\Program Files\euspell\bin32 (program directory)
c:\Users\Public\Program Files\euspell\spell (lexicon directory)
c:\Users\Public\Program Files\euspell\spell (server directory)
c:\Users\Public\Program Files\euspell\spell (user history or personal directory)
c:\tmp

It is save to accept the default directories, but for your convenience you can browse for another CD-station or for other directory names, or you can just enter another directory name.

- If you install from an Internet zip-file, a virtual CD will be created during un-

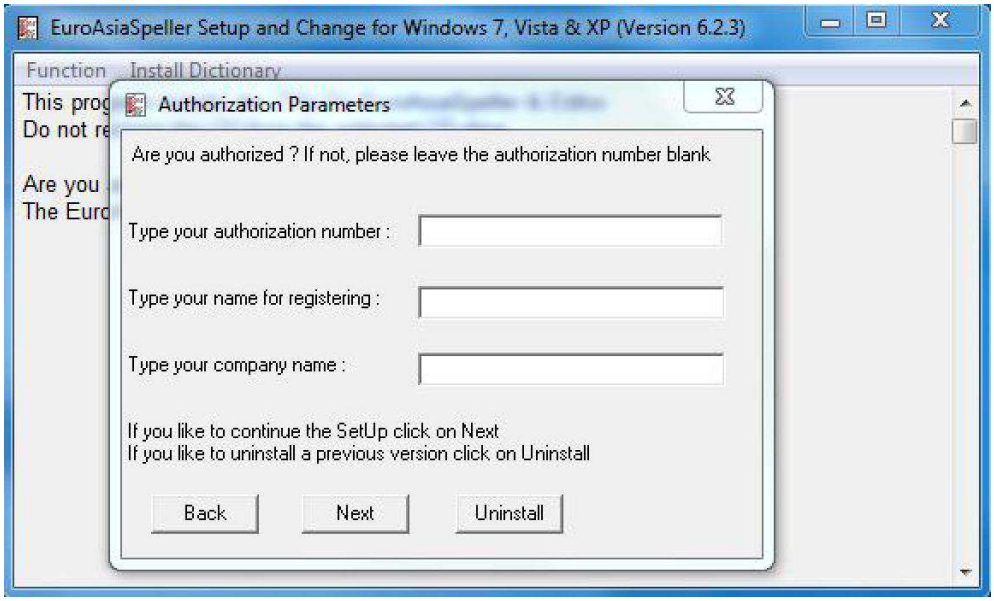
EuroAsiaSpeller Setup and Change for Windows 7, Vista & XP (Version 6.2.3)
Function
Install Dictionary
Authorization Parameters
Are you authorized ? If not, please leave the authorization number blank
Type your authorization number :
Type your name for registering :
Type your company name :
If you like to continue the SetUp click on Next
If you like to uninstall a previous version click on Uninstall
Back
Next
Uninstall

Fig. 3: The Authorization variables If you don't have an authorization number leave the authorization field of Dialogue empty! If you want to remove a previous installation of the EuroAsiaSpeller click on Uninstall. If an authorization key was supplied, it should specified in the authorization field before clicking Uninstall.

packing. The default path for this virtual CD is **c:\EuroAsiaSpeller**. This path is automatically recognized.

- The Program Directory "c:\Users\Public\Program Files\euspell\bin32" is the folder for executables: the EuroAsiaSpeller and the files needed to run the program.
- The Lexicon directory "c:\Users\Public\Program Files\euspell\spell" is the folder for all dictionaries
- The History Directory "c:\Users\Public\Program Files\euspell\spell" is the folder for the user history. You can also choose the current directory ".\".
  It is also possible to use a profile: "%USERPROFILE%\history". In this case the directory "history" is created in your personal folder which becomes: c:\Users\MyName\history If another user is also installing the speller his personal history will separated from your's.
- c:\tmp is a directory for temporal usage, probably already existing.

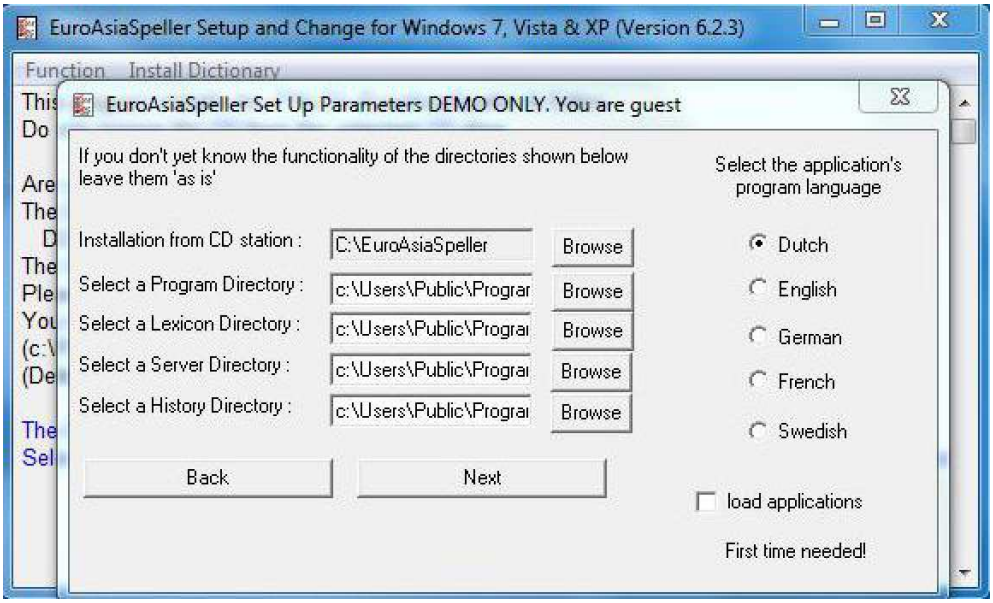

Fig. 4: The SetUp variables The above figure has a CD station on D: and has selected the default directories on drive C: to install programs and lexicons. If you like to browse to a different directory click on Browse. For a first installation select a program language (NL = Dutch, UK = English, D = German, F = French and SE is Swedish. Thereafter choose "Next load applications too". For loading dictionaries only choose "Next load dictionaries only" The loading of dictionaries is explained in fig. 9.

Note: you cannot assign the CD spell directory to the programs lexicon directory. So [CD]:\euspell\spell is invalid.

Subsequently, you can chose a program language version of the EuroAsiaSpeller: the Dutch (NL), English (UK), German (DE), French (F), or Swedish (SE) version. Thereafter you choose between "Next load applications too" (load executable programs) or "Next dictionaries only" to install new dictionaries.

If you have chosen for the installation of "Next load applications too", an optional group with shortcuts is created in either the "Start Menu" or its submenu "Programs". You can cancel this option if necessary.
You have to select OK to create a link for each application in the Euro Speller group either in Start or Programs Menu:

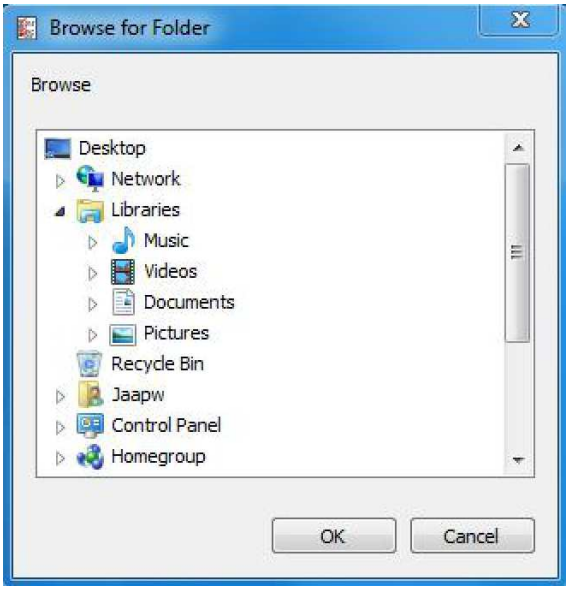

Fig. 6: Browsing for a directory

EuSpell, a text only version (reads text files) that is capable of handling Western and Central European, Baltic, Cyrillic, and Greek scripts, and complex scripts (Arabic, Hebrew, Hindi, etc.).

The menu "Install Dictionary" will be highlighted and should be used to select a target language. You now install the dictionaries one by one (see fig. 9 and 10). If you have a language license key you should enter this key in the edit field of the Change CD dialogue of fig. 10. This dialogue specifies the requested language and an edit field for the language key authorization number. Note this is not the general authorization number of fig. 3. The keys can be found in the accompanying form ACCESS KEYS FOR THE EUROASIASPELLER. If you don't have an authorization key a dialogue without an authorization field will be displayed. In order to install the dictionary you have to click on OK, otherwise click on CANCEL.
A correctly installed dictionary will be flagged in the menu giving you an overview of the dictionaries installed. If you try to re-install a demo dictionary the flag will be removed. You can re-install it by going back to the authorization dialogue and inserted a valid general authorization number. Re-installing the language will ask for the language authorization number. If a dictionary is not distributed on the current CD or in the extract of a zip-file the dictionary name is

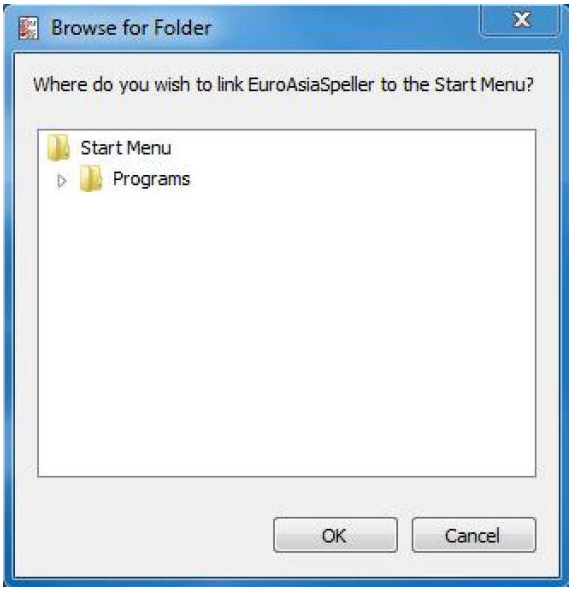

Fig. 7: The Start up menu, where to insert a start menu item. A group EuroAsiaSpeller will be inserted. Thereafter, the Desktop links (icons) will be inserted.

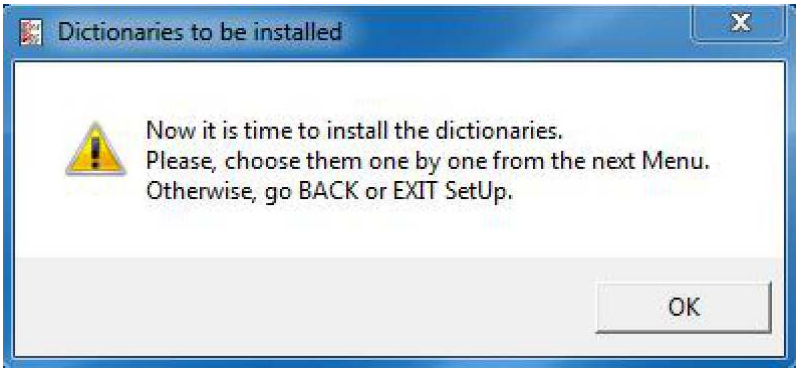
Dictionaries to be installed
Now it is time to install the dictionaries.
Please, choose them one by one from the next Menu.
Otherwise, go BACK or EXIT SetUp.
OK

Fig. 8: The Message notifying the installation of the dictionaries.

grayed.

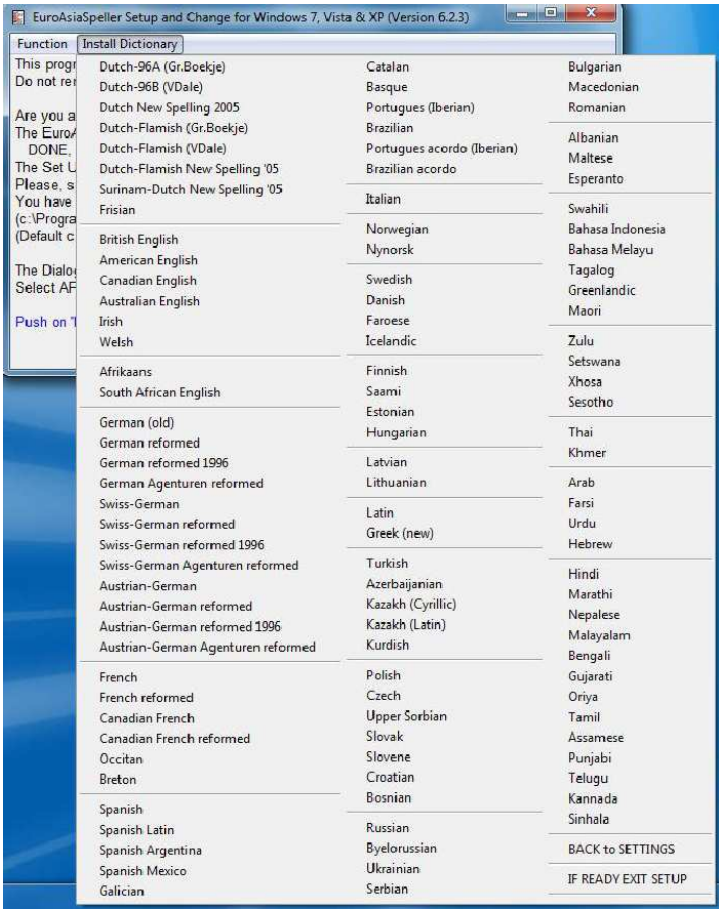

Fig. 9: If all dialogues are setup a language/dictionary can be select to be installed. The languages/dictionaries are grouped by the language family, that is Germanic languages, the Roman languages, etc. For the Slavic languages a latin-2 and a Cyrillic group exist. After the installation of one language a next one can be chosen. If ready you can Exit or go back to the Setting dialogue.

4. **HOW TO FINISH THE INSTALLATION**

After having installed all dictionaries, the installation should be finished by selecting FUNCTION | EXIT. The final dialogue "Up-date or Create Settings" appears. For the very first time you should always click on Yes to accept the current settings which you have chosen beforehand. If not initialized: the EuroAsi-

Fig. 10: Accept a demo Dictionary or Enter the language key code. If you are just trying the EuroAsiaSpeller for this language leave this field blank. If you didn't previously entered a general authorization number a dialogue without a number field is presented.

aSpeller will not run. If approved the installation procedure optionally adds the directory "c:\Program Files\euspell\bin32" to your search path in your "autoexec.bat" file (see the command PATH). If this file does not exist it will be created. For Windows Vista, Windows 7, and Windows 8 the "autoexec.bat" option does not apply.

4.1. **Usage**

After completing the installation you can use EuroAsiaSpeller.

The text version of the EuroAsiaSpeller is named "EUSpell.exe". The EuroAsiaSpeller reads UTF-8, UTF-16 and multiple Windows ascii files using codepages (WCP-1252, 1250, 1257, etc.). The file type to be saved can be chosen. For the CodePage type the spelling language determines the codepage.

(see the section "The EuroAsiaSpeller functions" pg. 15). Most text processing

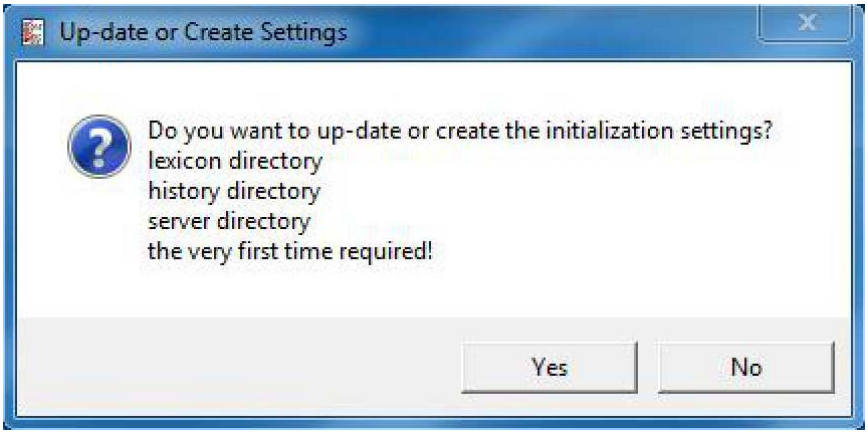

Fig. 11: Finishing the installation For the very first time you have to click on <Yes>. If the installer detects a new version of the library and executables you have to update the settings too. If you install new or updated dictionaries only this dialog is not shown.

packages can store and read their documents as Rich Text Format documents.

For Linux a wxWidgets richtext utility is available.

Both versions of the EuroAsiaSpeller can be started from the command line, or from the Start Menu.

For command line execution you use:

>euspell my_file.txt

You can also open files with the file open dialogue.

If you try to switch between languages an error message might be presented. Please, return to the Dutch (GB) language or one of the languages you have installed. Such a message can also occur when you use the speller for the first time. Please, select the Dutch (GB) language or one of the languages you have installed from the Settings Dialogue (evoked by the first button).

The EuroAsiaSpeller is easy to use. Its functionality is close to Microsoft's NotePad or WordPad, except for the additional spelling functions, the three buttons or menu items: settings, spelling and lexicon view.

4.2. **Fonts**

For most languages codepages are present in the default fonts such as *Times New Roman* and *Arial*. For Arabic, Hebrew and the languages of India Complex Scripts have to be enabled (Control Panel | Regional and Language Options).

4.3. **Limits**

The speller can only be used on Operating Systems which support Unicode (Windows XP, Windows Vista).
The EuroAsiaSpeller search function supports the use of the native language's citation marks.

5. **Languages available on the CD**

Afrikaans
Bahasa Indonesia
Bahasa Melayu
Basque
Bulgarian
Byelorussian
Catalan
Croatian
Czech
Danish (rettskrivning 2012)
Dutch (2005 (NGB))
Dutch (New Spelling (GB))
Dutch (New Spelling (VD))
Dutch: Flemish (2005(NGB))
Dutch: Flemish (New Spelling (VL/GB))
Dutch: Flemish (New Spelling (VL/VD))
Dutch: Suriname
English: Australian
English: Canadian
English: Irish Gaelic
English: New Zealandic
English: South Africa
English: UK

English: USA
English: Welsh
Estonian
Finnish
French
French (New Spelling)
French: Belgian
French: Belgian (New Spelling)
French: Canadian
French: Canadian (New Spelling)
Frisian
Friulian (Italy)
Galician
German: Austrian DPA
German: Austrian Reformed 1996
German: Austrian Reformed 2006
German: Austrian Traditional
German: DPA
German: Reformed 1996
German: Reformed 2006
German: Traditional
German: Swiss DPA
German: Swiss Reformed 1996
German: Swiss Reformed 2006
German: Swiss Traditional
Greek
Greenlandic
Hungarian
Icelandic
Italian
Latvian
Lëtzebuergesch (Luxembourg)
Lithuanian
Macedonian
Maltese
Maori
Norwegian: Bokmal
Norwegian: Nynorsk
Polish
Portuguese (+Acordo Ortográfico)

Portuguese: Brazilian (+Acordo Ortográfico)
Romanian
Russian
Saami
Serbian
Setswana
Slovak
Slovenian
Spanish
Spanish: Castilian
Spanish: Argentine
Spanish: Mexican
Spanish: Latin Am.
Swahili
Swedish
Tagalog
Turkish
Ukrainian
Vietnamese
Xhosa
Zulu
Setswana
Sesotho
Arabic
Hebrew
Persian/Farsi
Urdu
Azerbaijanian (Azari)
Kurdish (Northern)
Hindi
Marathi
Nepalese
Malayalam
Bengali
Gujarati
Tamil
Punjabi
Sinhala (Sri Lanka)
(see also fig. 9, see also "http://www.talo.nl/").
(the status of the languages is available from: "http://www.talo.nl/")

An authorized version of the EuroAsiaSpeller will be supplied on a CD and can be ordered from *TALŌ's address. This CD will include the package of languages ordered with the EuroAsiaSpeller.

Recent information about other languages can be obtained from http://www.talo.nl/" download menu.

6. **The EuroAsiaSpeller functions**

The EuroAsiaSpeller has been installed and one or more lexicons are ready to be used. For some languages an additional true-type font have to be installed. Free fonts from the Internet can be found on the CD. Please, use the Control Panel Item Fonts to install additional fonts.

The main spelling function consists of the three dialogues presented in fig. 12 and 13.

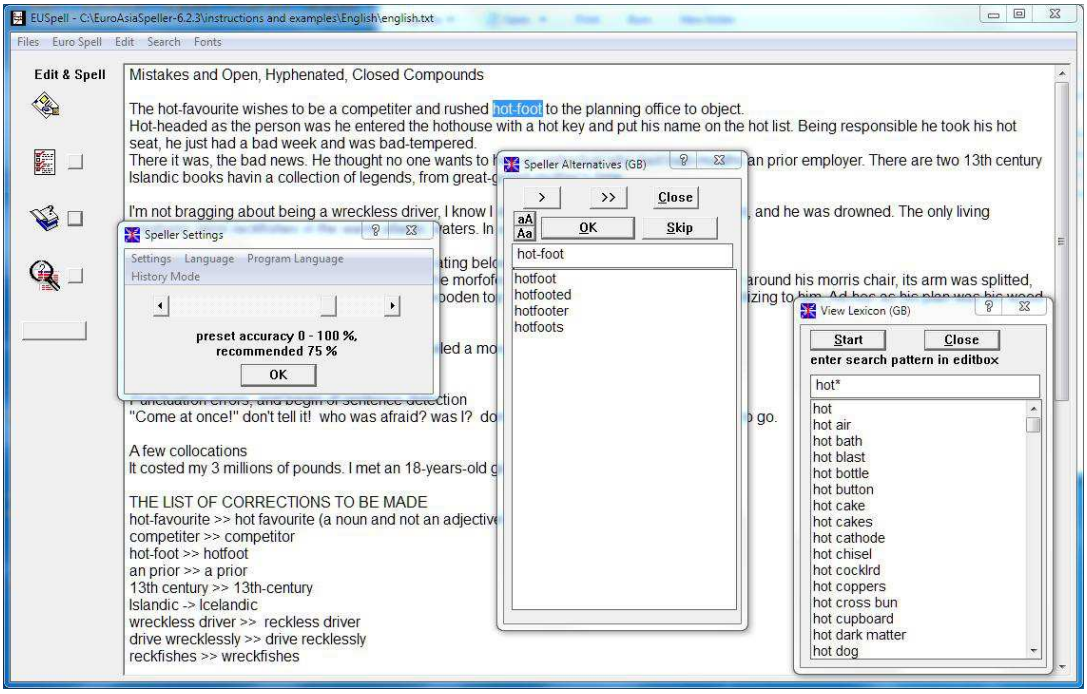
EUSpell - C:\EuroAsiaSpeller-6.2.3\instructions and examples\English\english.txt
Files Euro Spell Edit Search Fonts
Edit & Spell
Mistakes and Open, Hyphenated, Closed Compounds
The hot-favourite wishes to be a competiter and rushed hot-foot to the planning office to object.
Hot-headed as the person was he entered the hothouse with a hot key and put his name on the hot list. Being responsible he took his hot seat, he just had a bad week and was bad-tempered.
Speller Alternatives (GB)
Close
OK
Skip
hot-foot
hotfoot
hotfooted
hotfooter
hotfoots
Speller Settings
Settings Language Program Language
History Mode
preset accuracy 0 - 100 %,
recommended 75 %
OK
View Lexicon (GB)
Start
Close
enter search pattern in editbox
hot*
hot
hot air
hot bath
hot blast
hot bottle
hot button
hot cake
hot cakes
hot cathode
hot chisel
hot cocklrd
hot coppers
hot cross bun
hot cupboard
hot dark matter
hot dog
"Come at once!" don't tell it! who was afraid? was I?
A few collocations
It costed my 3 millions of pounds. I met an 18-years-old
THE LIST OF CORRECTIONS TO BE MADE
hot-favourite >> hot favourite (a noun and not an adjective
competiter >> competitor
hot-foot >> hotfoot
an prior >> a prior
13th century >> 13th-century
Islandic -> Icelandic
wreckless driver >> reckless driver
drive wrecklessly >> drive recklessly
reckfishes >> wreckfishes

Fig. 12: The EuroAsiaSpeller functions (text version) with the dialogs Select settings, Speller alternatives and Lexicon view opened

The functionality of the dialogues is identical for EUSpell and any other application using the EuroAsiaSpeller design.

Each of these dialogues is evoked by a button press at the left side of the main window or by selecting a menu item under Euro Spell.
For EUSpell from top to bottom: the settings dialogue, the alternative dialogue and the lexicon view dialogue. The Settings Dialogue can also be evoked with the function key F10, the alternative dialogue with F11 and the lexicon view dia-

logue with F12. The dialogues can be opened concurrently. Moreover manual editing during spelling is allowed. The functionality of the three buttons can also be evoked from the menu item "Euro Spell".

One or more files are opened with "File | Open". A fourth button at the left of the main window becomes significant when multiple files are selected. This button is used to spell the next file and will display the word "next" as long as there are files to spell.

The use of the three buttons "Settings", "Alternatives" and "Lexicon view" are specified below:

6.1. **Settings**

Text • Digits/Text ✓ • OCR
The user can set the type of text to be spelt to "text" only, "digits/text" or "OCR". Additionally abbreviations and punctuation control can be enabled.

Enable abbreviations ✓
Each abbreviation can be checked for a match with the abbreviation list. Abbreviations should be enabled when the punctuation mechanism checks capitalization after full stops (periods).

Double word check ✓
Internet Address check ✓
Checked to History
Checked from History
Yes
Auto ✓
No
Respell only
For control purposes the user can enable warning history messages. Items can be send **to** the history or/and corrections can be executed **from** the history data base.
Messages **to** the history concern a check whether the user is certain to include an item in the history.
Messages **from** concern the conditions of using an item for correction of the text. The user can have full control over these manipulations. For "Checked from History" — the correction — the user can choose between:

"Yes", full warning, all history items will get a message box [Y/N],

"Auto", only warnings for the latest respelling pairs and those which have

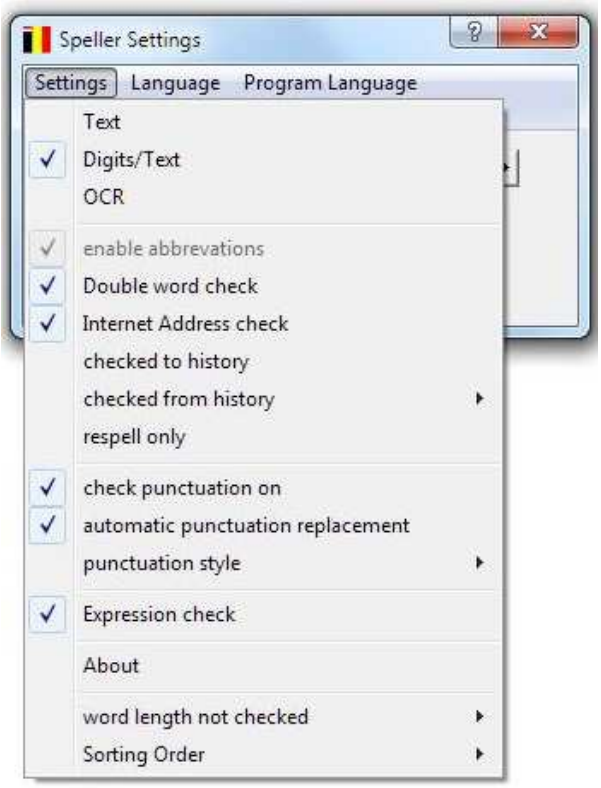

Fig. 13: The History menu functions: always [Y/N] query, automatic learning of the history and default no query (automatic correction).

not been used frequently will evoke a message box (the server history is considered to be certain to be used automatically),

"No", no warnings at all, except for those history items which have got a warning mark.

The "Respell only" item only uses re-spelling pairs from the server, style and user history. Normal spelling is omitted. This mode can be used for text modifications between British and American English orthographies.

Check punctuation on ✓
Automatic punctuation replacement
Punctuation style
For control purposes the speller can send warning punctuation messages to the regular alternative box, for punctuation errors a single correct pattern is dis-

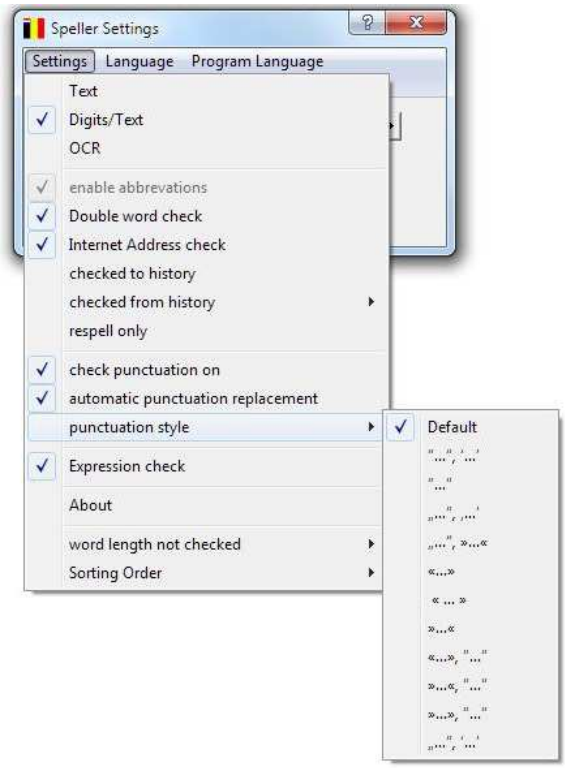

Fig. 14: The 11 different citation styles used during punctuation checks.

played only. This pattern can be accepted or the user can change the punctuation pattern manually. For English the punctuation marks "..." are used (style 1), for Dutch „..." (style 3) and for German „..." (style 4). The default punctuations of a language are chosen with style 0 (see fig. 14). For the Scandinavian countries the proper guillemet characters are applied. These guillemets differ per Nordic country. Type writer quotation marks are converted to graphical marks. Greek, Hebrew, and Arabic should use the default style (0) only.
The patterns which belong to the punctuation marks can be modified by editing the "xxx-xxx-xxx.punc" files in the spell directory (c:\euspell\spell) (xxx-xxx-xxx should be exchanged by the current language name, see the list of language names). The ansi text editor EUspell can control punctuation which runs over to the next line. Here, a paragraph boundary in the form of a carriage return is assumed to be present. However, many texts use carriage returns to close each line of the text and there is no intention to associate returns with a paragraph boundary. To prevent the paragraph assumption the Euro Spell menu's item

"*Paragraph mode*" should disabled.
The punctuation control principle also checks for missing capital characters at the start of sentences, this form of checking occurs after a dot, question mark or an exclamation. For this type of correction the control of abbreviations should be enabled and private abbreviations should be added to abbreviation list (for English the file eng-gbr-std.abbr in the spell directory; the spell directory was entered in fig. 3).
MicroSoft's functionality often exchange single and double quotes into there ascii equivalents " or '. The punctuation mechanism can exchange the type writer quotation marks into graphical quotations. The later ones will be printed and are written to file, but keep the above in mind! EUspell keeps any single and double quote "as they are".

The preset punctuation style is defined in the punctuation data base. However, the national language style can be changed in another style using different kind of quotes (see also the Appendices B and C).

Expressions or collocations
The expression function enables spelling of multiple words, usually named collocations. Examples are Geographical Names (Black Sea, Snake River), lexical or grammatical collocation (a house, a union, not, an house, a union). Expressions or collocations have to be defined in advance. For English, Dutch, French and German a broad set of expressions has been included. The user can add his own expressions to the personal or user history file (see also Appendix D).

Paragraph mode [on/off] (main Menu | Euro Spell)
The "Paragraph mode" state in EUSpell's main speller menu can be set to enable punctuation checks at the end of each text line, and/or multiple word spelling (grammatical and lexical collocations and expressions. Each return is considered as the end of a paragraph and therefore the next paragraph should start with a capital character. This check is disabled when "Paragraph mode" is switch off.

Word length not checked
The user can set the word length to be disregarded.

Sorting Order
The user can switch between an alphabetic sorting order or an sorting order based on similarity. Similarity is relative to an user made mistake and most similar item is put in top of the list.

6.2. **Language**

The language menu item shows the list of languages which are installed. Each new language has to be installed in advance by the SetUp function. Thereafter it will appear in the language menu. Dictionaries cannot be copied from the CD, but should be installed using the SetUp script on the EuroAsiaSpeller CD.

6.3. **Program language**

The program language menu sets the language of the menu texts.

6.4. **History**

The default settings for the history should be "save correction as in text". This mode is case sensitive.
See also "History and respelling" below.
History items are presented in an International Dialog Message Box. Therefore, history messages are clearly different from regular spell check operations.

6.5. **Alternatives**

The Alternative button or menu item either activates spelling from the cursor or spelling of the selected story. If the speller stops for an error the alternatives are displayed. The user can select an alternative and press the "replace" button. He can also "skip" the item. If there are not enough alternatives the user can press the ">" or ">>" button and select the proper alternative. Upper and lower case can be toggled by the "T" button. Finally the dialogue is closed by pressing the "Close" button. During spelling the user can return to the text and make small changes to the text, e.g., remove superfluous spaces. To continue spelling, the Alternative dialogue should be activated by the "replace" or "skip" button.
If a word to be spelled's orthography is changed the change is stored to the User History (see below). If a word to be spelled's orthography is accepted as a correct word the word is put in the User History too. The above storage can be modified in two ways:

a) keeping the Control Key down to store the item for the current Session Only. The button text is placed between brackets: [...], equivalent to "Skip All" and "Change All".

b) keeping the Alt Key down to attach a History Warning signal to assign the

item as ambivalent. The button change text is appended with an asterisk (*), e.g. "Change *".
The Linux RichText XML spellchecker use the CONTROL+ALT Key combination to warn against an ambivalent change of orthography.

6.6. **Lexicon view**

The lexicon view dialogue enables the user to view the lexicon with wild cards. A pattern is specified in the upper edit field and the "start" button is pressed search in the lexicon for matching words. The dialogue is closed by pressing the "Close" button.

6.7. **A spelling session**

A spelling session is defined as the action of spelling the current text using the current language. If no text is selected the file is spelled from the cursor position. If a "section of the text" is selected the lines belonging to the selected text are only spelt. A spelling session is finished after checking every word or after closing the dialogue (see also "Storage of user correction actions").

6.8. **History and respelling**

The history is an intelligent correction system which consists of pairs of words (or pairs of expressions) and words normally assigned to the user's dictionary. The history is divided into a protected section and a user section. For languages with recent reformed dictionaries the protected history consist of a collection of binary respelling items (see Appendix D). For languages with special problems the protected history also includes grammatical and lexical collocations. Moreover, special attention is given to capitalization of geographical expressions and national institute names. The user section is updated during spelling, the protected history can only be edited afterwards. New items can be appended at the end of the protected history list. Items in the user section can gain a higher priority on being used. Items which are not used anymore are subject for deletion if space is needed for new items. The server items are stored in the file "xxx-xxx-xxx-cdic", however, this file should be created and edited in advance. The style information is stored in the file "xxx-xxx-xxx.styl" and the user data are stored in the file "xxx-xxx-xxx.pdic" (see below). These files can be found in the "\euspell\spell" folder. This folder is set during installation.
The history mode can be set to:
save as lower case: The pairs of words are converted to lower case and the case in the text is significant for correction. This mode is meant for nearly all lan-

guages except German.
save correction as in text: The correction section of the pairs of words is saved as in the text but the mistakes are converted to lower case. Therefore case is not significant for mistakes. This mode is meant for German, but can be used for any language.
save all as in text: The pairs of words are saved as in the text. Case is significant. This mode is meant to collect raw information.

6.9. **Storage of user correction actions**

The buttons "OK" and "Replace/Change" are used either for an acceptance of a word (a correct word) or a correction of the text chosen by the user. The actions are stored in the history system and are re-used each time the same correct words is found in the text and a similar correction should be made. Depending on the speller's setting corrections can be made automatically or approved using the confirmation dialogue [Y/N].
The user can also chose to apply an action to the above buttons for the current session and current language only, and not to apply them for any future session. Such a state is forced by keeping the CTRL Key of the keyboard pressed. The text of buttons will be replaced by *[OK]* and *[Replace]* (see fig. 15). If brackets appear none of the temporary history records will be stored or remembered for later usage.

The user can add a history warning signal to the Replace/Change button by holding the ALT-key (for Linux hold ALT+CTRL-key). The button's text becomes *Replace* *, where the asterisk means "add a warning" (see fig. 15).
The other functionality follows standard principles found in other applications.

6.10. **The Linux EuroAsiaSpeller example**

The Linux version of the EuroAsiaSpeller behaves similar to the Windows version. The application is a variant of the wxWidgets RichText example. The dialogues are slightly different but their functionality is similar.
The dialogues are shown in fig. 16. The SettingDialog presents all setting within one windows, while the Windows version applied a menu driven function. The LexiconViewDialog lists a range of comparable words for the search pattern "hot*". This dialogue also shows expressions, or word combinations, in this case a few compounds such as "hot air, hot blast, hot cake" are shown which call for a space. The SpellCheckDialog also focuses on these word combinations, but the uncertainty was solved from the history, an already known respell-

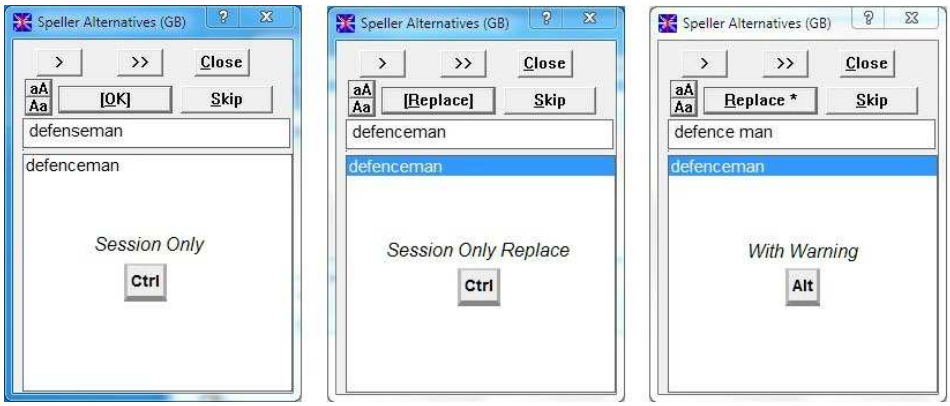

Fig. 15.: HotKeys and the representation of Session Only and SpotHistoryWarnings, resp. Session Only Added Words (Control HotKey), Session Only Replacements (Control HotKey), and Replacements with an Always Warning (Alt HotKey).

ing pair (see the message box "Replacements").

For Linux the CTRL key puts the OK and Replace button between brackets too.

Please do not use complex scripts. WxWidgets is not yet capable to process them.

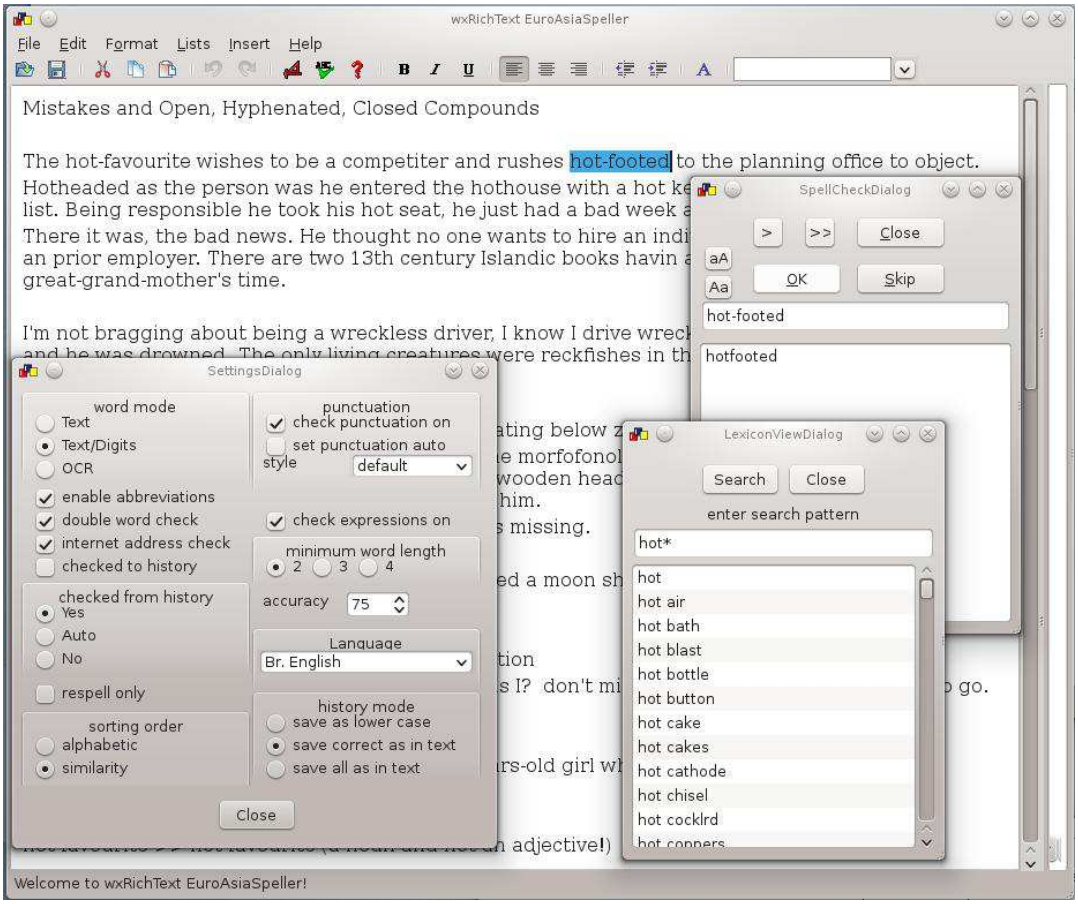

Fig. 16: The Linux version of the EuroAsiaSpeller

7. **APPENDICES**

Appendix A: List of short name language identifiers used in the Windows preset dialogs.

Dutch new spelling 2005	NGB		
Dutch Flemish new spelling 2005	NVB		
Surinam Dutch new spelling 2005	SUR		
Dutch GB	NLA	British English	GB
Dutch VD	NLB	American English	USA
Dutch Flemish GB	NLG	Canadian English	CGB

Dutch Flemish GVD	NLD	Afrikaans	AFR
German	D	Swiss-German	CH
German new spelling	D2	Swiss-German new spelling	CH2
Austrian German	AU	Austrian German new spell.	AU2
French	F	Canadian French	CF
French 1990	F2	Canadian French 1990	CF2
Occitan	OCC		
Spanish Peninsular	E	Spanish Argentine	ARG
Spanish Mexican	MEX	Spanish Latin Am.	LAM
Catalan	CAT	Galician	GAL
Basque	EUS	Portuguese	P
Brazilian	BZL	Italian	I
Swedish	S	Danish	DK
Norwegian	N	Nynorsk	NY
Finnish	SF	Russian	RUS
S.Afr.English	SAE	Australian English	AUS
Estonian	EST	Latvian	LAT
Lithuanian	LT	Polish	PL
Czech	CZ	Slovak	SLK
Slovene	SLV	Croatian	CRO
Serbian	SRB	Albanian	ALB
Romanian	ROM	Macedonian	MAC
Latin	LTN	Esperanto	ESP
Saami	SAM	Swahile	SWA
Frisian	FRK	Greenlandic	GRN
Maltese	MLT	Tagalog	TAG
Ukrainian	UKR	Greek	GR
Bah.Melayu	MAL	Bah.Indonesia	BID
Zulu	ZUL	Xhosa	XHO
Sesotho	SOT	Setswana	SET
Azerbaijanian	AZR		
Arabic	ARA	Hebrew	HEB
Persian/Farsi	FAS	Urdu	URD
Hindi	HIN	Marathi	MAR
Nepalese	NEP	Kurdish (N.)	KUR
Malayalam	MLM	Bengali	BNG
Gujarati	GJR	Tamil	TML
Punjabi	PJB	Sinhala	SNL
Friulian	FUR	Luxembourgish	LTZ

The national flag is showed in dialog's left top corner.

Appendix B: List of language base names (lexicons and abbreviation files)

Language	base name	short name	language id
Dutch New Spelling (GB)	dut-nld-grb	NLA	318 (obsolate)
Dutch New Spelling (VD)	dut-nld-vda	NLB	319 (obsolate)
Dutch 2005 (NGB)	dut-nld-vda	NGB	481 ‡
Flemish New Sp.(VL/GB)	dut-fle-grb	NVG	300 (obsolate)
Flemish New Sp.(VL/VD)	dut-fle-vda	NVD	301 (obsolate)
Flemish 2005 (NGB)	dut-fle-ngb	NVN	482 ‡
Surinam Dutch (SUR)	dut-sur-std	SUR	480 ‡
British English	eng-gbr-std	GB	302
American English	eng-usa-std	USA	303
Canadian English	eng-can-std	CGB	420
Australian English	eng-aus-std	CGB	439
Irish Gaelic	eng-gle-std	GLE	470
German	deu-ger-old	D	304
German new spelling	deu-ger-new	D2	324
German 1996 spelling	deu-ger-n96	D96	492 (1996,obsolate)
German new agenturen	deu-ger-dpa	D3	435
Swiss German	deu-swi-old	CH	305
Swiss German New	deu-swi-new	CH2	365
Swiss German 1996	deu-swi-n96	CH9	493 (1996, obsolate)
Swiss German Agenturen	deu-swi-dpa	CH3	436
Austrian German	deu-aut-old	AU	431
Austrian German New	deu-aut-new	AU2	432
Austrian German 1996	deu-aut-n96	AU9	494 (1996,obsolate)
Austrian German Agent.	deu-aut-dpa	AU3	437
French	fra-fre-std	F	306
French new spelling	fra-fre-ref	F2	349
Canadian French	fra-can-std	CF	421
Canadian French new spelling	fra-can-ref	CF2	422
Spanish	spa-spa-std	E	307
Spanish Latin	spa-lam-std	LAM	515
Spanish Argentina	spa-arg-std	ARG	516
Spanish Mexican	spa-mex-std	MEX	517
Catalan	spa-cat-std	CAT	308
Italian	ita-ita-std	I	309
Friulian	ita-fur-std	I	518
Portuguese	por-por-std	P	310
Brazilian Portug.	por-bra-std	BZL	311

Portuguese acordo	por-pac-std	PAC	496
Brazilian acordo	por-bac-std	BAC	497
Swedish	swe-swe-std	S	312
Danish	dan-dan-std	DK	313
Norwegian	nor-nob-std	N	314
Nynorsk	nor-nno-std	NY	315
Finnish	fin-fin-std	SF	316
Estonian	est-est-std	EST	260
Latvian	lav-lav-std	LAT	261
Lithuanian	lit-lit-std	LT	371
Iceland	isl-isl-std	ICE	322
Greek	grc-ell-std	GR	317
Turkish	tur-tur-std	TR	370
Hungarian	hun-hun-std	H	372
Polish	pol-pol-std	PL	374
Czech	ces-ces-std	CZ	375
Slovak	slk-slk-std	SLK	378
Russian	rus-rus-std	RUS	379
Ukrainian	ukr-ukr-std	UKR	423
Bulgarian	bul-bul-std	BLG	424
Afrikaans	afr-afr-std	AFR	427
South African English	afr-eng-std	SAE	433
Latin	lat-lat-std	LTN	434
Frisian	dut-fry-std	FRK	438
Euskara (Basque)	spa-eus-std	EUS	451
Faroese	fao-fao-std	FAR	452
Galician	spa-glg-std	GAL	453
Saami	nor-sme-std	SAM	454
Romanian	ron-ron-std	ROM	457
Albanian	alb-alb-std	ALB	458
Macedonian	mkd-mkd-std	MAC	459
Bahasa Indonesia	idn-ind-std	BID	460
Greenland	kal-kal-std	GRN	461
Croatian	hrv-hrv-std	CRO	456
Serbian	srp-srp-std	SRB	462
Bahasa Melayu	msa-msa-std	MAL	463
Slovene	slv-slv-std	SRB	455
Tagalog	tgl-tgl-std	SRB	464
Swahili	swa-swa-std	SWA	439
Maltese	mlt-mlt-std	MLT	466

Esperanto	epo-epo-std	ESP	467
Thai	tha-tha-std	TH	468
Byelarussian	bel-bel-std	BLR	469
Welsh	eng-wel-std	WEL	473
Maori	mao-mao-std	MAO	474
Azerbaijanian	azr-azr-std	AZR	475
Armenian	arm-arm-std	ARM	476
Georgian	geo-geo-std	GEO	477
Zulu	zul-zul-std	ZUL	478
Vietnamese	vie-vie-std	VIE	479
Xhosa	afr-xho-std	XHO	483
Setswana	afr-set-std	SET	484
Sesotho	afr-sot-std	SOT	495
Hindi	ind-hin-std	HIN	485
Arabic	ara-ara-std	ARA	486
Hebrew	heb-heb-std	HEB	487
Marathi	ind-mar-std	MAR	488
Persian/Farsi	per-fas-std	FAS	489
Urdu	urd-urd-std	URD	490
Nepalese	nep-nep-std	NEP	498
Kurdish (Northern)	kur-kur-std	KUR	499
Malayalam	ind-mal-std	MLM	500
Bengali	ind-ben-std	BNG	501
Gujarati	ind-guj-std	GJR	502
Tamil	ind-tam-std	TML	504
Punjabi	ind-pan-std	PJB	507
Sinhala	lka-snl-std	SNL	510
Luxembourgish	ltz-lux-std	LTZ	520

Appendix C: For English the following files exist:

index lexicon	eng-gbr-std.indx
main lexicon	eng-gbr-std.lexi
abbreviation list	eng-gbr-std.abbr (ascii, utf8, utf16)
stop list	eng-gbr-std.stop (ascii, utf8, utf16) to be created by the user group)
server history	eng-gbr-std.cdic (ascii, utf8, utf16, to be created by the user group)
system style guide	eng-gbr-std.styl (supplied)
user history	eng-gbr-std.pdic (ascii, utf8, utf16, created automatically by the user)
punctuation control	eng-gbr-std.punc (ascii, utf8, utf16)

List of default directories

Program dirrectory	c:\euspell\bin32
Lexicon directory	c:\euspell\spell
Server directory	c:\euspell\spell
History directory	c:\euspell\spell (user history only)
Temporary directory	c:\tmp

The maxima

abbreviations	18,000 bytes
protected + user history	38,000 pairs (mean word length 25/20 bytes, plus 15.000 collocations)
punctuations	1,000 pairs (mean length per item 10 bytes)
stop list	18,000 bytes

Appendix D: The format of the history.

The history concept is a part of the speller engine with is capable to correct text automatically. New items can be send to the history during spelling too. If the user accepts a respelling case (a correction) or an unknown word by clicking OK the item is copied to the history. Single words can only be copied automatically. Expressions or collocations consist of more then one word. These strings have to be copied to the history afterwards, e.g., the collocation "due for an illness -> due to an illness" might be copied into the history file using the format presented below.
Any correction request based on a history record can be succeeded by a message box to confirm the correction [Y/N].

The records in history files (protected server history "xxx-xxx-xxx.cdic" and user history "xxx-xxx-xxx.pdic") have the following format (see fig. 17):

priority "correct word" "erroneous word" [flag]
300 transcription transcribtion
300 G_major g-major

Note that the underscore symbolizes a space and that the space is used as an argument separator. This underscore replacement is not used in the Speller Manager but only applies to the history files themselves.

The erroneous word also can be an erroneous expression such as:

300 an_even_greater even_a_greater

The above example evokes the correction of "an even greater" to "even a greater". A history record is an expression if the erroneous section of the record consists of multiple words! In the English language these expressions are also called collocations. In German they are also called "Redewendungen". In Dutch these expressions refer to the concept of "vaste uitdrukkingen".
Note that expressions can only be entered or modified with the Speller Manager tool.

The flag, a special signal

The *, + and ! character can be put after a respelling pair to evoke a deliberated warning. A * at the end of a record always evokes a warning message box, even for automatic correction. A ! at the end of a record only warns for initial capitalizations in the text. This means that lower case strings can be corrected without intervention of the user if "form history" warnings are disabled. A + at

```
300 graveller graveler
300 gravelled graveled
300 gravelling graveling
300 teetotallers teetotalers
300 Walloon Wallonian
300 Walloons Wallonians
300 Walloons Walonian
300 Walloons Walonians
300 accoutre accouter
300 accoutrement accouterment
300 advertise advertize
300 advertised advertized
300 advertisement advertizement
300 advertisements advertizements
300 advertiser advertizer
300 advertisers advertizers
300 advertises advertizes
300 advertising advertizing
300 advice advices
300 adze adz
300 ad_hoc ad-hoc
300 aeroplane airplane
300 aeroplanes airplanes
300 aesthete esthete
300 aesthetic esthetic
```

Fig. 17: An example of the English history. The first entry respells ***graveler → graveller.***

the end of a record only warns for initial minuscules in the text. This means that upper case strings can be corrected without intervention of the user if "from history" warnings are disabled. These signals are meant to be used for those cases where conflicts exist. E.g., according to the new German orthography the German word "schwermachen" (with verbs) now should be divided into two words: "schwer machen" But the noun case "das Schwermachen" should not be respelled! The following record prevents automatic conversion for Capitals:

300 schwer_machen schwermachen +

For Dutch the spelling reform changed the spelling of "paardekoper" in "paardenkoper" but "Paardekoper" can be a person's name too.

300 paardenkoper paardekoper + (lower case in text)

only corrects automatically for lower case strings.
The next definition

300 paardenkoper paardekoper ! (upper case in text)

would only correct for an initial upper case letter

However, the next case ALWAYS warns!

300 paardenkoper paardekoper *

These warning can be attached to both single words and expressions (multiple words).

Capitalization and the history format
For **save as lower case** words stored in lower case do accepts lower case words in the text and its first letter capitalization during spelling, e.g., the entry **behaviour** accepts both words **behaviour** and **Behaviour**.
An entry with a first capitalized letter is an incorrect format! If first letter capitalization is needed this capitalization has to be entered as a respelling pair in the history file, e.g.,
washington → **Washington.**
This pair also accepts first letter capitalized words in the text.

The same principles applies to collocation/expression pairs, e.g., the pair:
african congress → **African-Concress**
respells the lower case to upper case, but it also respells
african Congress or **African Congress** in the text to **African-Congress.**
So the **save as lower case** mode is non-specific in terms of error detection of case.

For **save correction as in text** the text is spelled and the word is accepted as correct if the same case was entered in the history. At future occasions the identical case is only accepted. The entry **Washington** differs from the entry **washington.** The correction pair
washington → **Washington**
corrects but accepts the new first letter capitalized word too. The same applies to collocations/expressions, e.g. the pair: African Congress → African-Concress only applied to the exact case. For other cases entries have to be repeated.

Add your own alternatives to user mistakes instead of ours

It is possible to define your own alternatives in the History records or to get alternatives to a word of high importance. The format of these records have a preceding period in the second column:

```
300 ._word_to_be_warned_for
300 ._suggestion1,suggestion2 mistake_word
300 ._suggestion1,suggestion2 mistake_collocation
```

In the first case (a period only) similar cases to the "word_to_be_warned_for" are given In the second case (a period plus suggestions) the suggestions are supplied, and in the third case multiple word collocations are used. No wild cards are allowed! (using the Speller Manager the underlines should be replaced by a space and the special records are preceded by ". ".

Wild cards in the expressions

For expressions wild cards are allowed, but should be used with great care.

```
A "in expression" * matches up to 7 multiple words
A "in expression" $ matches a single words
A "in expression" ? matches a single character
A "in expression" + matches a single lower case character
A "in expression" | matches a single upper case character
A "in expression" # matches a single number like the number "456"
A "in expression" ^ matches a single character number like "6"
```

The record :

```
300 EURO_# euro_#
```

converts lower case to upper case for all numerical combinations after the word EURO.

Note that an * wild card cannot be used at the end of an expression!

The history can be edited using the Speller Manager application (see fig. 18a & b). This application automatically locates the "default history". Strings can be entered "as is" and no underline is needed to simulate a space.

Abbreviations in upper case only

To convert lower case abbreviations to upper case you have to add the following type of definition in the history file ("eng-usa-std.cdic" or "eng-usa-std.pdic")

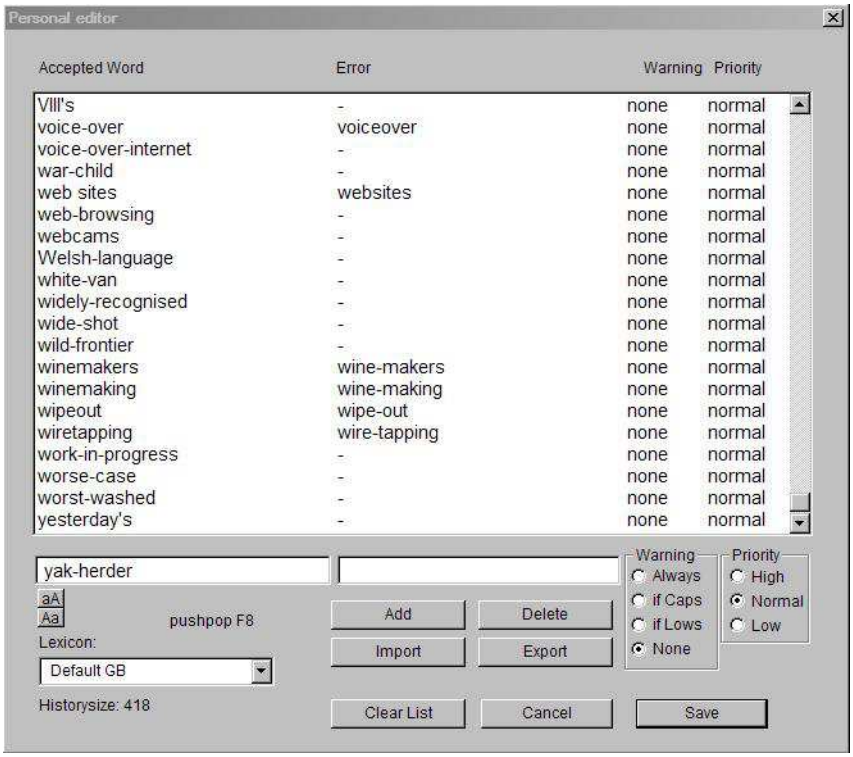

Fig. 18a: An example of the Speller Manager application using the personal history for the Br. English language. In the example an item is set ready to be processed, either by editing or by a push-pop transfer. You can push an item to the server history but you have to click the target window (server) before transfering the item. This is a protection mechanism to have the correct target windows. Please note, there also exists an abbreviation window!

300 IBS ibs

For regular inclusion of words both lower and upper case characters are accepted.

Protected and user history

The user history is automatically maintained during usage. The protected history is read-only and can only be altered by editing. For most languages the protected history includes predefined respelling pairs. These predefined cases cannot be read and the unreadable section should not be altered. The range not to altered is the section between the @_his and @_end statements. Any new item

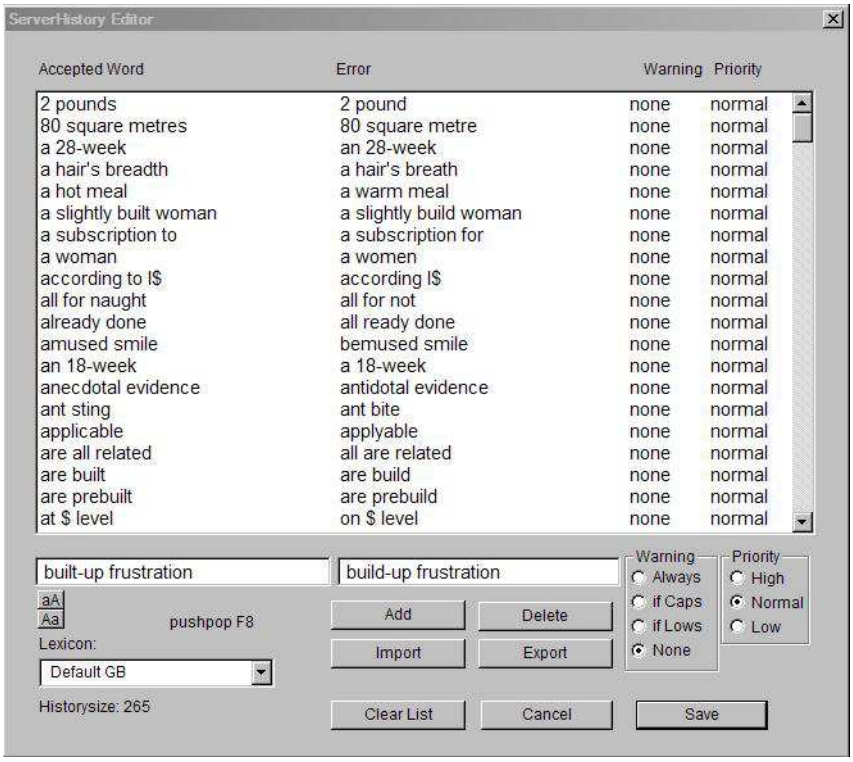

Fig. 18b: An example of the Speller Manager application using the server history for the Br. English language. In the example an item is set ready to be processed, either by editing or by a push-pop transfer. You can push an item to the personal history but you have to click the target window (personal) before transfering the item. This is a protection mechanism to have the correct target windows. Please note, there also exists an abbreviation window!

should be entered before the @_his statement or after the @_end statement in the protected history file. If any predefined item does not fit, the item can be redefined in the very top of the file (above the @_his statement).

Special issues of the EuroAsiaSpeller

The EuroAsiaSpeller engine is capable of handling Western European, Central European, Baltic, Cyrillic, Greek texts and so is the utility EUSpell. This utility also handles complex scripts such as Arabic and Hindi.

For additional technical info please, request for the document "talo_s_lib6236.pdf"

We hope that you will enjoy the EuroAsiaSpeller.

*TALŌ bv
Dr.J.C.Woestenburg
e-mail jaapw@talo.nl
tel +31 35 69 32 801
tel +31 65 46 83 544
fax +31 35 69 75 993